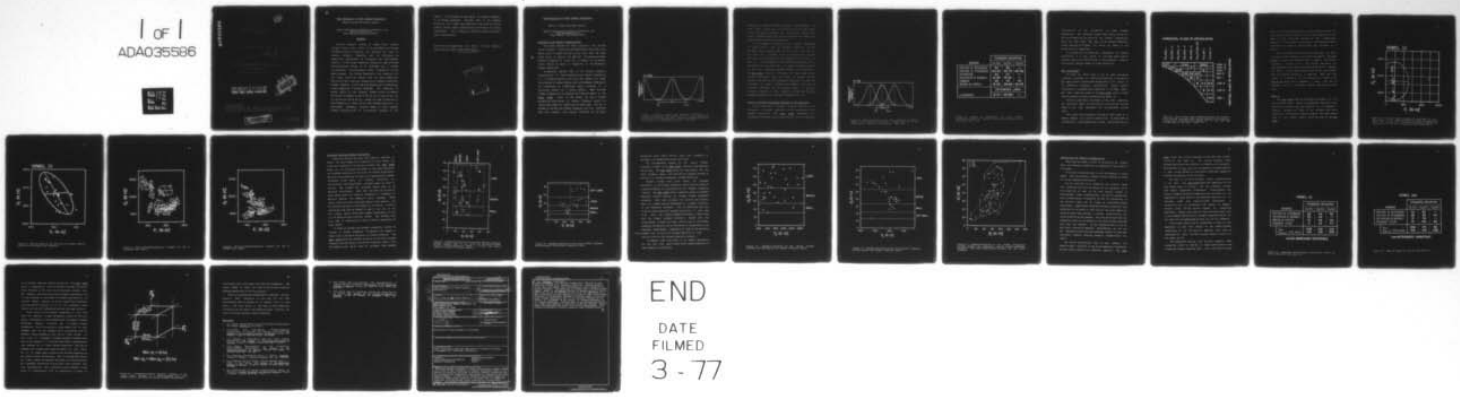
AD-A035 586 SPEECH COMMUNICATIONS RESEARCH LAB INC SANTA BARBARA CALIF F/G 5/7
SOME STATISTICS ON VOWEL FORMANT VARIABILITY.(U)
NOV 76 D J BROAD, H WAKITA N00014-76-C-0483

UNCLASSIFIED NL

1 OF 1
ADA035586



END

DATE
FILMED
3 - 77

AD A035586

(12)

Some Statistics on Vowel Formant Variability*

David J. Broad and Hisaghi Sakita

see 1473

Speech Communications Research Laboratory, Inc.
800A Miramonte Drive
Santa Barbara, California 93109

DDC
REPRODUCED
FEB 14 1977
C

COPY AVAILABLE TO DDC DOES NOT
PERMIT FULLY LEGIBLE PRODUCTION

---

*Paper presented at the American Association of Phonetic Sciences, November 15, 1976, San Diego, California. Parts deleted from the spoken version are in parentheses.

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

Copy available to DDC does not permit fully legible reproduction

Some Statistics on Vowel Formant Variability

David J. Broad and Hisashi Wakita

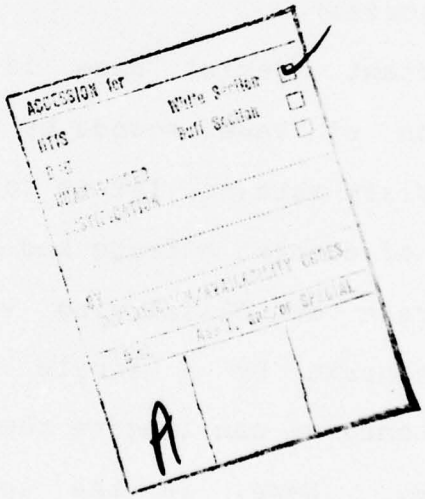
Speech Communications Research Laboratory, Inc.
800A Miramonte Drive
Santa Barbara, California 93109

ABSTRACT

Acoustic phonetic studies of vowels often present average results with little of the information on formant variability which is essential for defining the sizes of acoustic phonetic categories and for evaluating the comparative significance of contextual and inter-speaker effects. In this study, phonetics instruction tapes provided 828 steady-state tokens of 17 unrounded and 13 rounded non-nasalized, non-retroflexed vowels produced by a single female speaker. The formant frequencies were measured via the linear prediction method, and the most steady-state 108.8 ms of each token was located by an automatic algorithm. Formant analysis errors and wild samples were eliminated by visual inspection of scatter diagrams. The remaining 779 tokens (94.7% of the original tokens) result in standard deviations of between 13 and 87 Hz for F_1; 32 and 145 Hz for F_2; and 28 and 159 Hz for F_3. Except for some indications of multi-modality for cases of very high standard deviation, no discernable systematic relation seems to exist between formant variability and (1) articulatory features of the

vowels, (2) nativeness of the vowels, (3) formant frequency, or (4) formant bandwidth. Although some of the standard deviations are larger than expected on the basis of earlier studies, certain vowel configurations are found to be highly reproducable. This finding is related to acoustic phonetic vowel classification.

[This work was supported by the Office of Naval Research under Contract Number N00014-76-C-0483.]



Some Statistics on Vowel Formant Variability

David J. Broad and Hisashi Wakita

Speech Communications Research Laboratory, Inc.
800A Miramonte Drive
Santa Barbara, California 93109

Variability and Phonetic Classification

(Our study concerns the basic question:) how refined should phonetic categories be? Measurements show that every speech event is unique and that no one ever says the same thing twice in exactly the same way. We therefore define phonetic categories to allow for a degree of variation, (which should be small in comparison to the difference between categories).

An important special case is the acoustic phonetic classification of vowel sounds by their formant frequencies [1]. To simplify matters, let us for the moment set aside the problems of context effects and inter-speaker differences and concentrate on classifying vowels produced in a controlled context by a single speaker. Then for some formant frequency we can imagine the situation shown in the first slide. Here, in the center, we see a normal probability distribution of formant frequency values for (some large number of) repetitions of some vowel. Now let us attempt to divide this formant frequency axis into intervals that will, together with similar divisions of the other

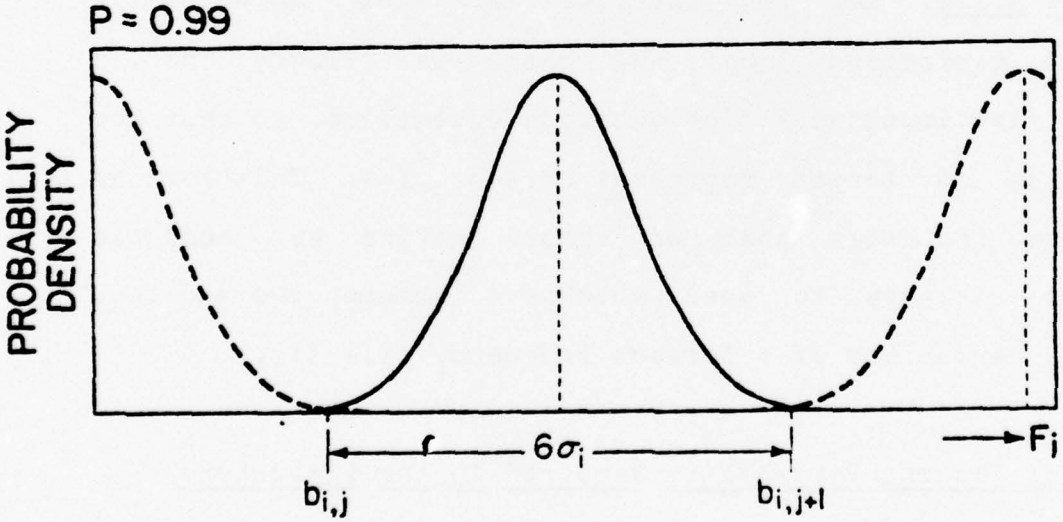


Figure 1. (Center) Hypothetical Caussian distribution of formant frequencies for repetitions of a vowel under fixed conditions by a single speaker. Phonetic category boundaries are spaced at six standard deviations. From [1].

formant axes, define the vowel categories acoustically. It makes sense to make such a division according to the spread of the distribution, because the distribution defines the range of formant values that can be considered to result from samples drawn from the same category.

In this example an interval of six standard deviations is marked off to show one possible definition for the allowable intra-category range for this vowel. This may be too large for a category, for it would imply that a sample falling at the boundary was a rare example of the sound, as well as a rare example of the neighboring category (with a probability of less than 0.01). It might therefore be better to make the categories somewhat smaller, as illustrated in the next slide. Here the intervals are only about 2-1/2 standard deviations long, and there is sizable overlap between distributions for neighboring categories, so that the boundaries no longer represent rare samples. This sort of reasoning indicates that we should define an acoustic phonetic category to span somewhere between two and four standard deviations of a formant frequency axis [1].

Review of Formant Variability Reported in the Literature

In this approach to defining acoustic phonetic vowel categories it becomes important to know the values of the standard deviations. The next slide summarizes the information available from various sources in the literature

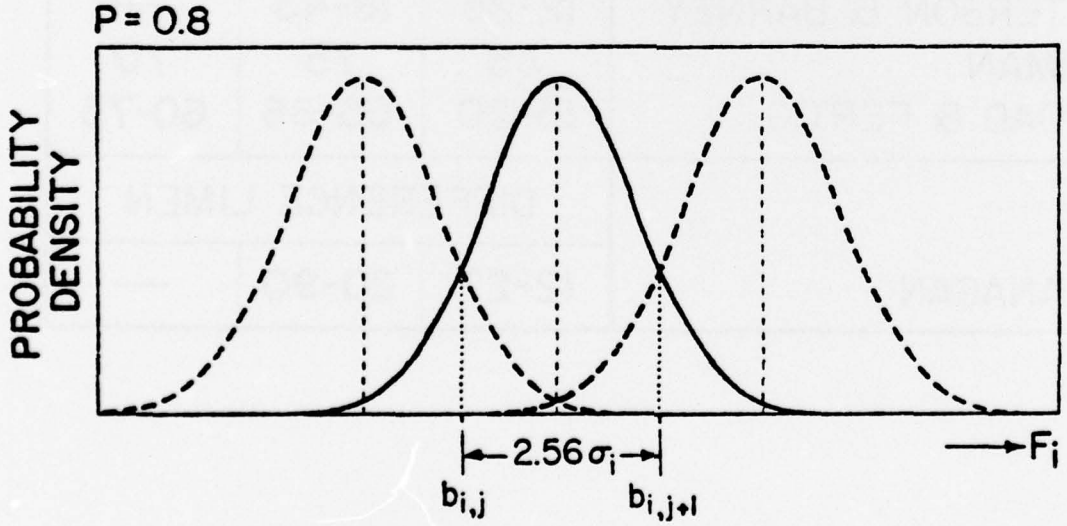


Figure 2. Same as Figure 1, but with category boundaries spaced 2.56 standard deviations. Eighty percent of the distribution is within the boundaries. From [1].

SOURCE	STANDARD DEVIATION		
	σ_{F_1}(Hz)	σ_{F_2}(Hz)	σ_{F_3}(Hz)
POTTER & PETERSON	20	95	—
POTTER & STEINBERG	20-40	40-70	60-90
PETERSON	20	27	—
PETERSON & BARNEY	12-25	18-43	—
ÖHMAN	35	75	70
BROAD & FERTIG	15-20	55-66	60-75
	DIFFERENCE LIMEN		
FLANAGAN	12-27	20-90	—

Figure 3. Summary of information on vowel formant variability taken or inferred from various sources in the literature [2-6, 8-9].

[2,4,5,6,8,9] on the variability of vowel formant frequencies. The different studies show similar values for the inter-repetition variation of the formant frequencies; these are also about the same as the formant difference limena reported by Flanagan [3], which are shown in the bottom line for comparison.

Our purpose was to extend this information on formant variability to a rich subset of the vowel space, and to overcome some of the difficulties of spectrographic analysis with digital analysis based on linear prediction.

Data and Methods

Our data are taken from a set of tape recordings originally produced as instruction materials in phonetics. The tapes contain steady-state productions of 823 tokens of 30 different non-nasal non-retroflex voiced vowels, (including 17 unrounded vowel types and 13 rounded types). The vowel types are shown in the next slide, which is taken from the phonetic theory of Peterson and Shoup [7].

Within a given cell, the symbol on the left represents the unrounded vowel for the place of articulation, and the symbol on the right represents the corresponding rounded vowel.

The vowels were originally recorded on audio tape by a female speaker (in a sound treated room). We digitized the recordings at 10,000 samples per second, and stored them on

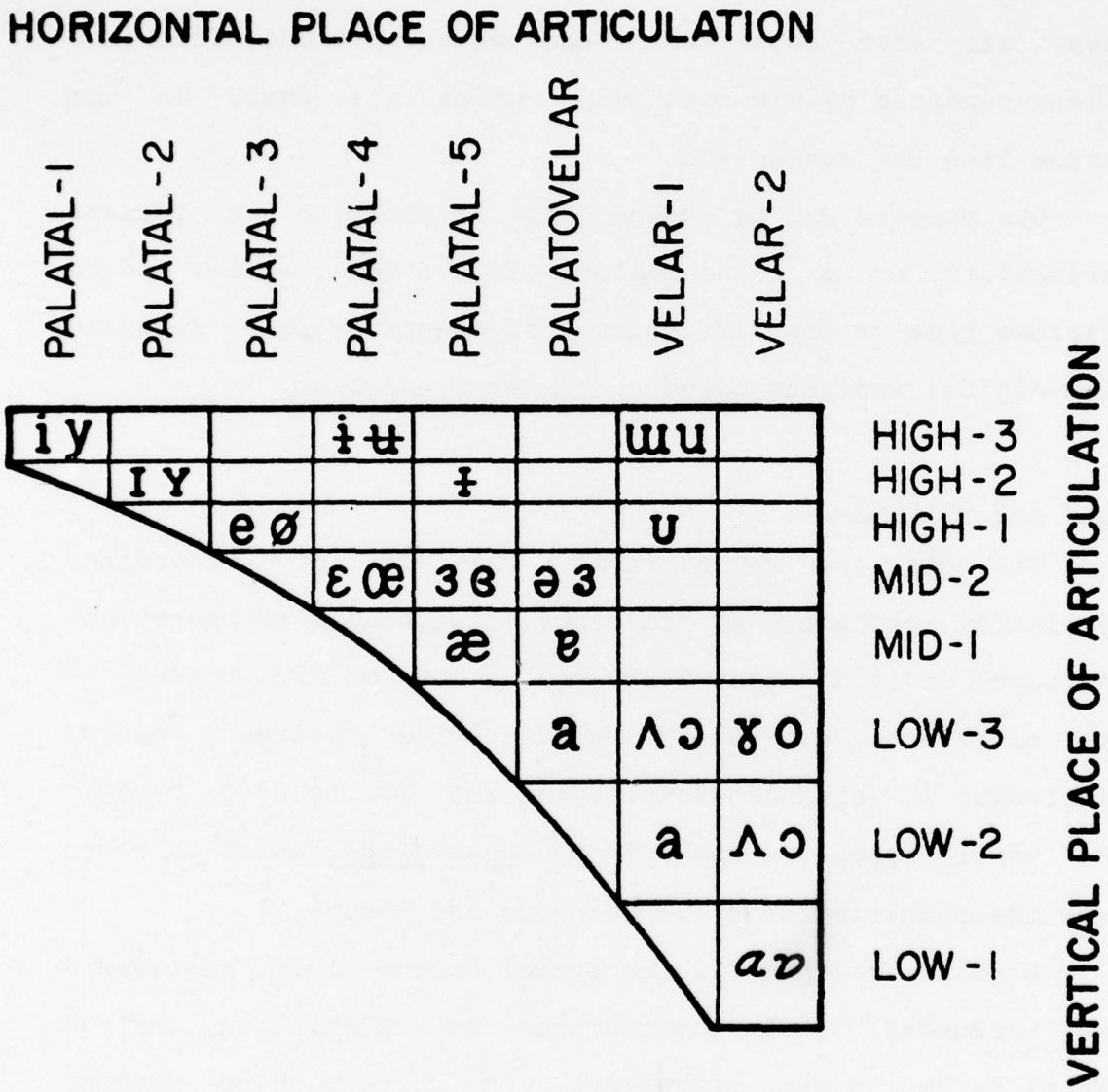


Figure 4. The 30 vowel types studied organized on a phonetic chart showing horizontal and vertical places of articulation. In each cell, unrounded vowel symbols are to the left and rounded ones to the right. After [7].

disk. Formant Frequencies were estimated (every 19.2 ms) via the linear prediction method (using a frame length of 32 ms), then a simple algorithm located the most steady-state sequence of five frames of each vowel. The average of these five frames was taken to characterize the utterance as a whole.

Next, the results were plotted for each vowel in order to locate formant analysis errors and wild samples. The next slide illustrates our procedure with a typical distribution including a formant error. The ellipse represents the 2-standard-deviation region calculated from all the samples; the effect of the single formant error on the first formant mean and standard deviation is apparent. When the wild sample is eliminated, the distribution is as shown in the next slide, which has an expanded F_1 scale. Now the 2σ ellipse provides a reasonable general description of the data.

Results

The next slide shows the resultant 2σ ellipses for all the unrounded vowels. There is considerable variation in the size of the ellipses, and there seems to be no systematic relation between the sizes of the ellipses and their locations on the formant frequency diagram. The same remarks hold for the rounded vowels, which are shown in the next slide.

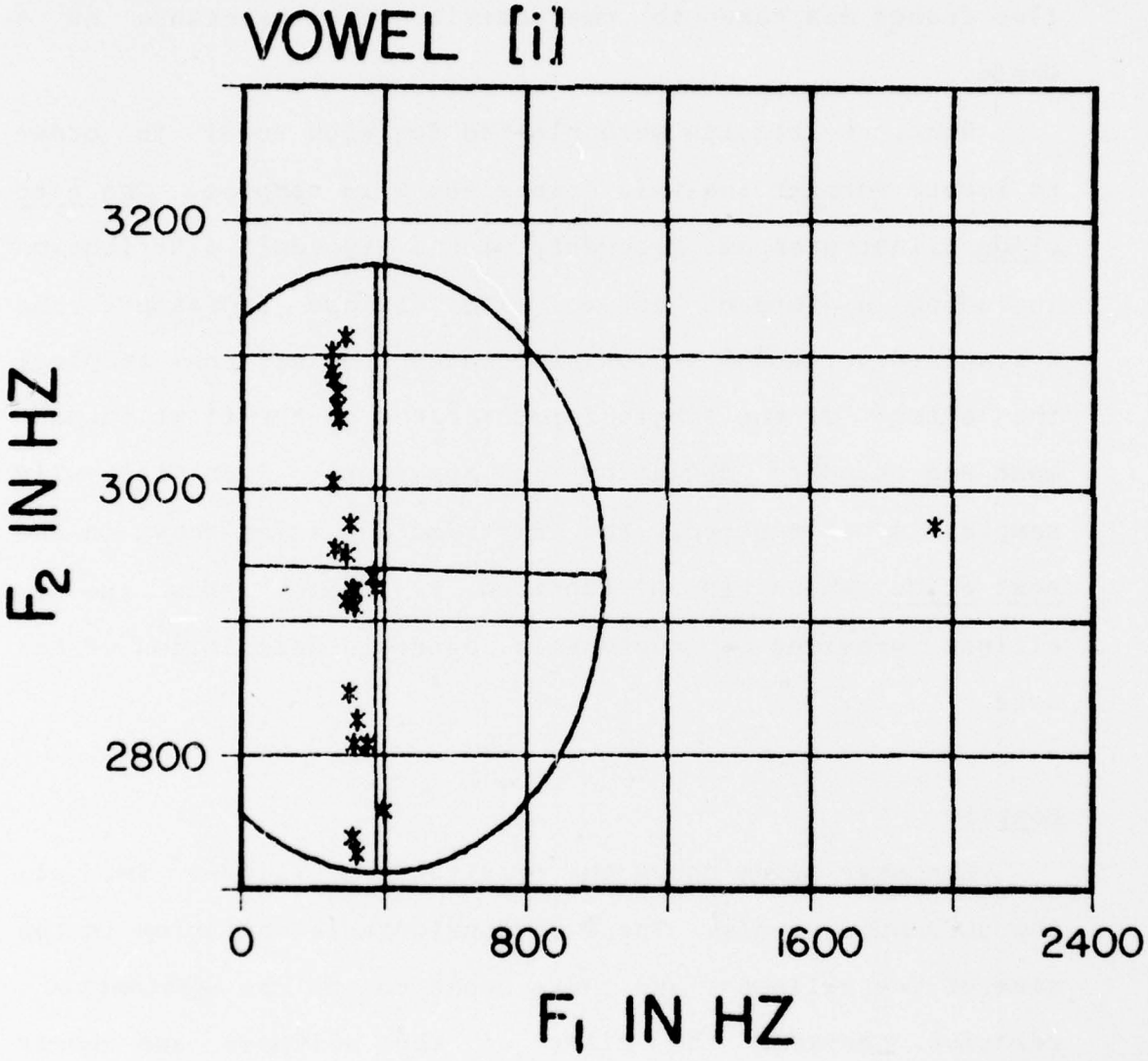


Figure 5. First two formant frequencies measured for [i]. The sample at the right represents a formant identification error. The ellipse is two standard deviations in radius, the intersection of its axes is the mean of the distribution.

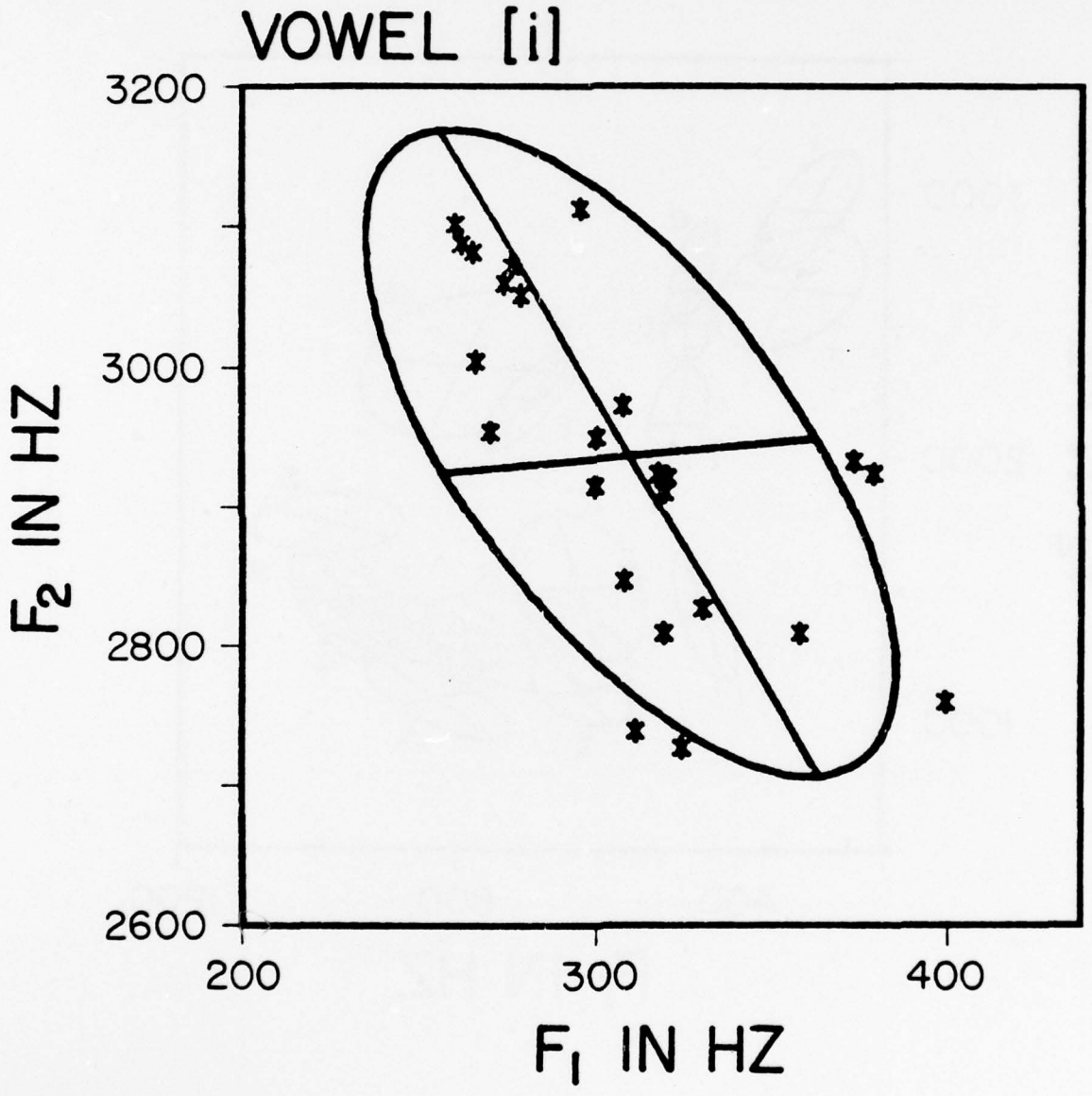


Figure 6. Same as Figure 5, but with the wild sample removed and the first formant scale expanded.

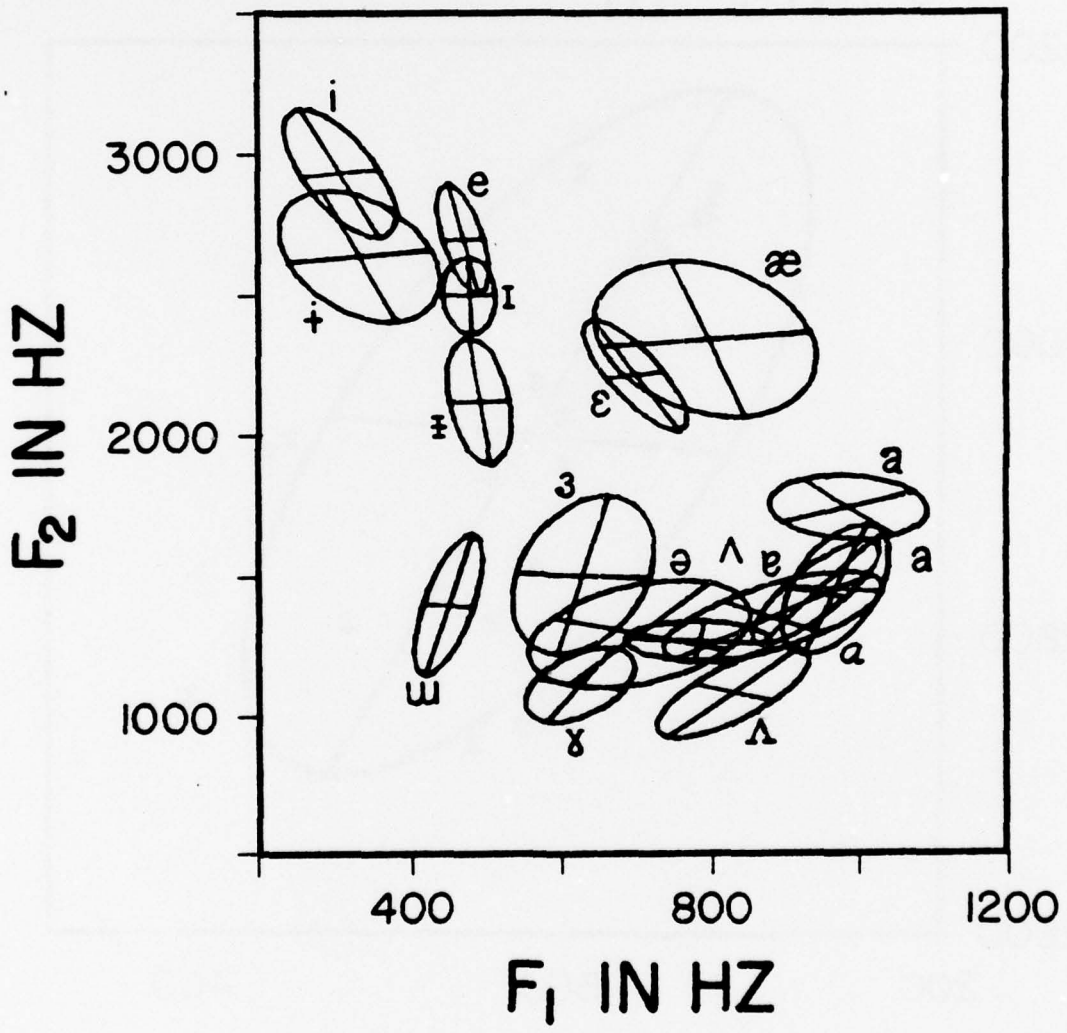


Figure 7. The two-standard-deviation ellipses for the 17 unrounded vowel types.

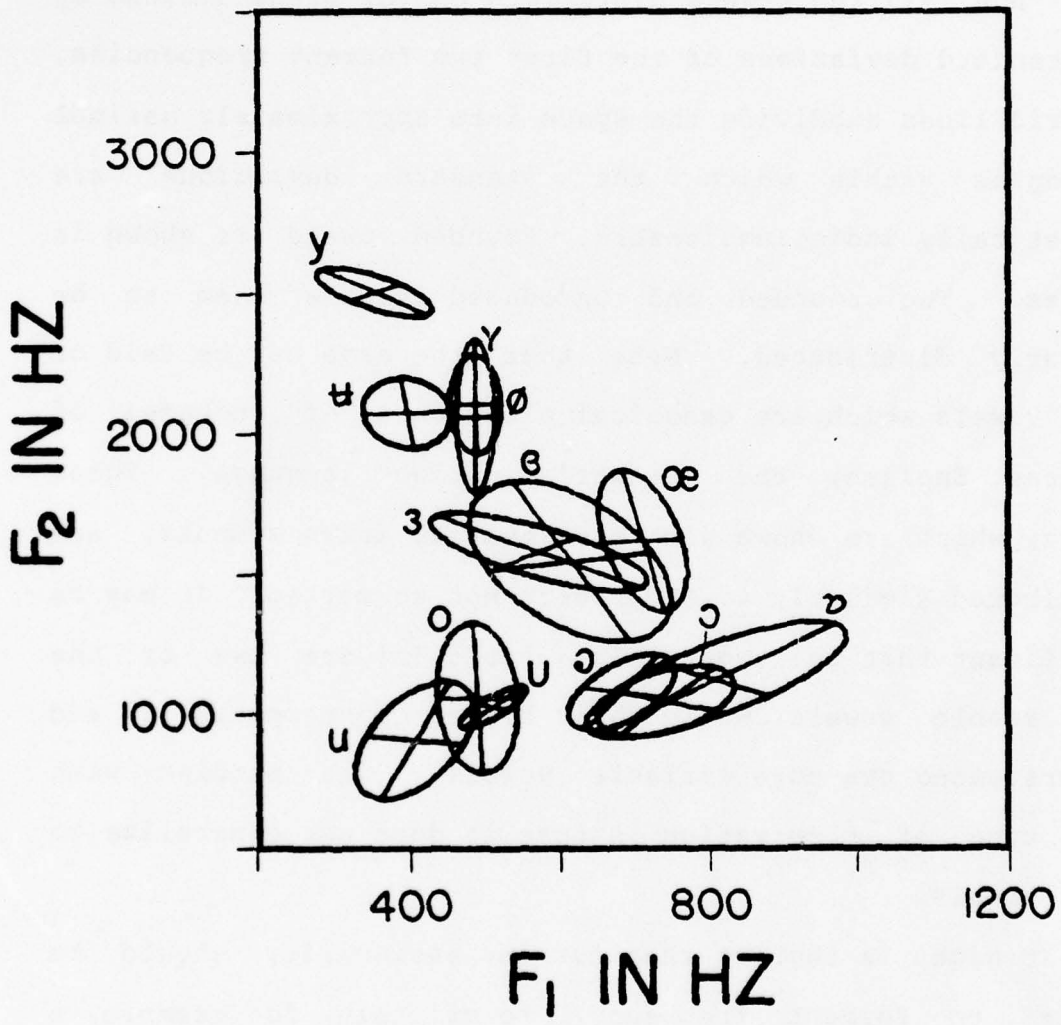


Figure 8. The two-standard-deviation ellipses for the 13 rounded vowel types.

Variables Affecting Formant Variability

(There are several variables that might be expected to affect the inter-repetition variability of vowel tokens.) To study the variability a little more closely, the next slide shows how all the vowels distribute on the plane defined by the standard deviations of the first two formant frequencies. The grid lines subdivide the space into approximately maximal rectangles within which the standard deviations are statistically indistinguishable. Rounded vowels are shown in circles. The rounded and unrounded vowels seem to be similarly distributed. Note that the same can be said of those vowels which are canonical allophones of phonemes of American English, the speaker's native language. These vowels, which are shown with more heavily drawn symbols, are distributed similarly to the vowels not so marked. It may be significant that the two vowels [ɪ] and [ʊ] are two of the more stable vowels, while their higher counterparts [i] and [u] are among the more variable vowels. The problem with this type of observation is that it does not generalize to other vowels.

It might be thought that formant variability should be related to formant frequency, to maintain, for example, a fixed ratio of standard deviation to formant frequency. The next slide shows the standard deviation of the first formant frequency plotted against the formant frequency itself. Only the weakest trend can be seen: no extremely high standard

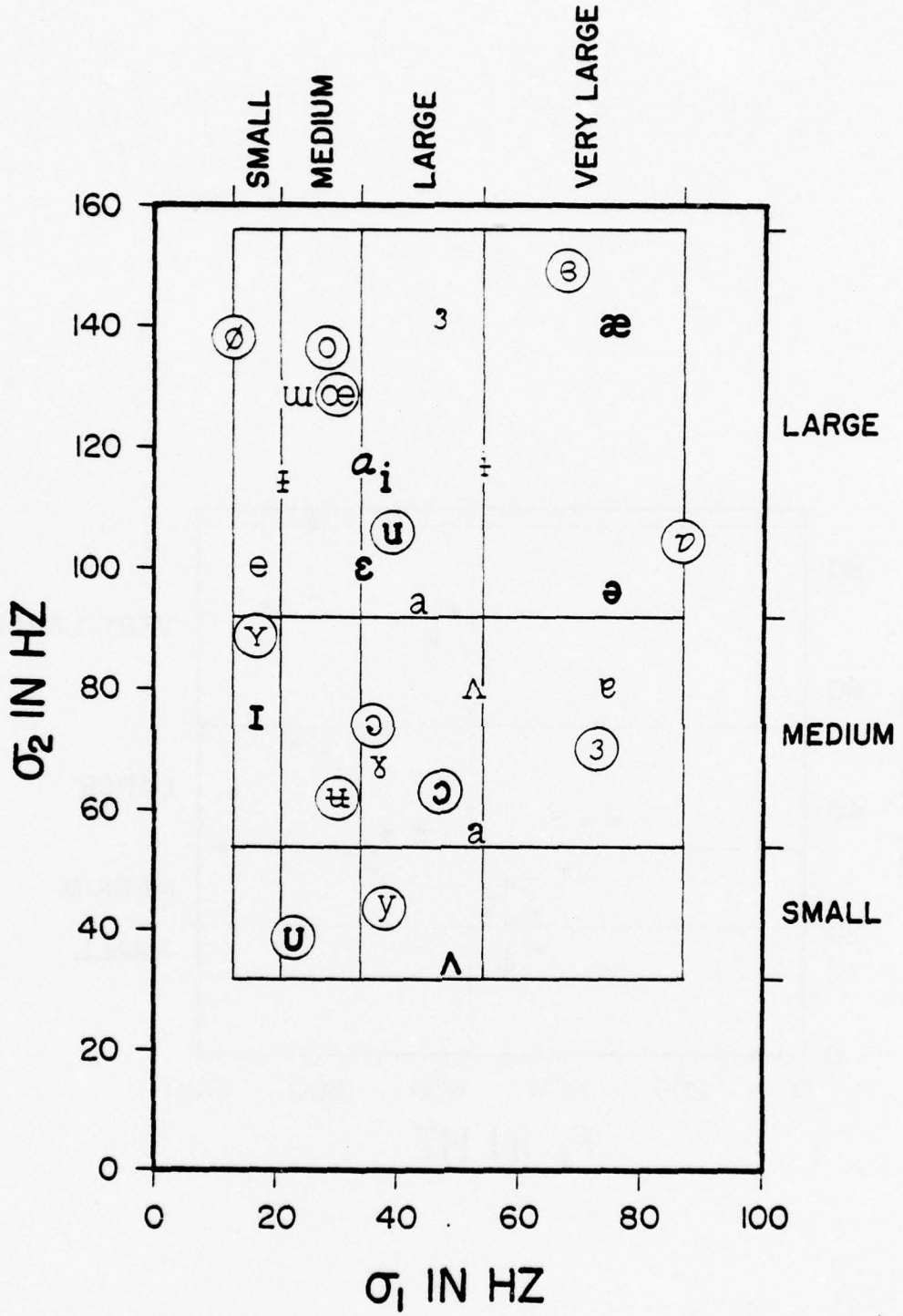


Figure 9. Standard deviation of the second formant frequency plotted against that of the first for the thirty vowels studied. Rounded vowels are shown in circles and native vowels are shown with darker symbols.

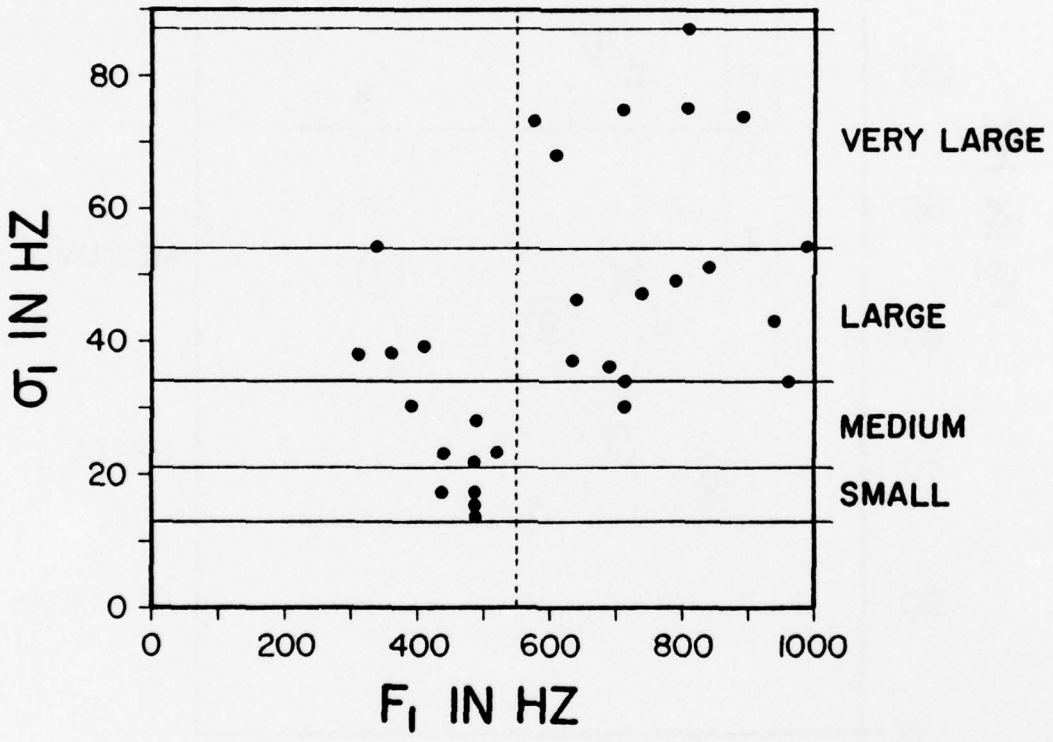


Figure 10. Standard deviation of the first formant frequency plotted against the formant frequency itself.

deviations exist below 550 Hz; above that frequency no extremely low standard deviations are found.

The corresponding diagram for the second formant frequency is shown in the next slide. Here no trend whatever can be seen. The next slide shows the relationship for the third formant; again, the distribution suggests nothing in the way of a frequency-dependent standard deviation.

Another variable that could affect the standard deviations is the formant bandwidth, which might be taken as a physical measure of uncertainty for the formant frequency location. The next slide shows a composite graph of standard deviation plotted against the average formant bandwidth estimates. Each loop encloses the data for one formant. There is a slight positive dependence of standard deviation on the bandwidth estimates - the six bandwidths less than 75 Hz are associated with standard deviations of less than 30 Hz. Also, the largest bandwidth estimate is associated with the largest standard deviation. Beyond this, the relationship is again very weak, and not much of the formant frequency variability can be attributed to uncertainty in the frequency measurement, (especially in view of the existence of one example ([ʉ]) in which the standard deviation is less than one fifth of the average bandwidth estimate.)

It appears, then, that there is no simple explanation for the fact that these vowels differ among themselves in their degree of variability.

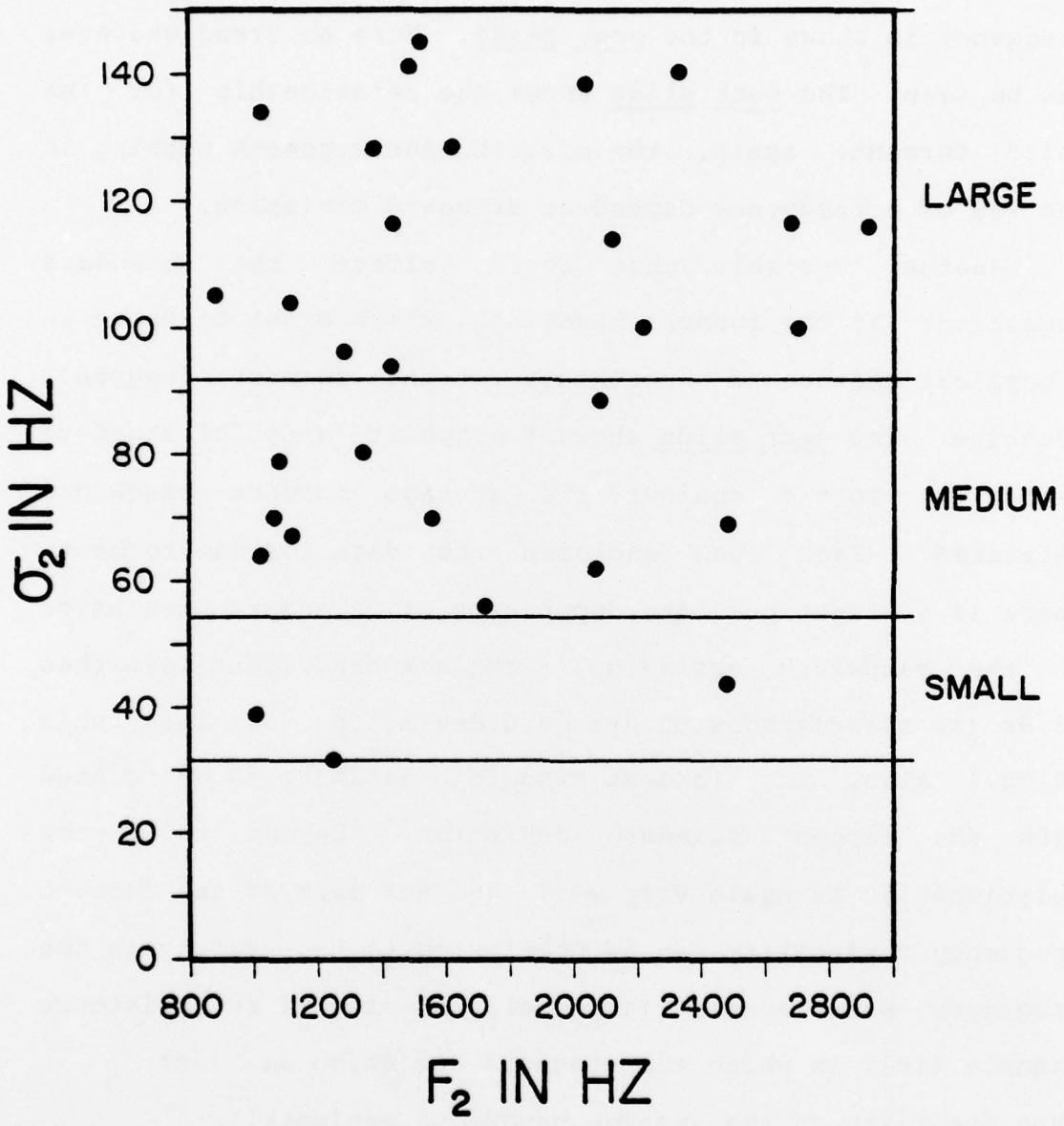


Figure 11. Standard deviation of the second formant frequency plotted against the formant frequency itself.

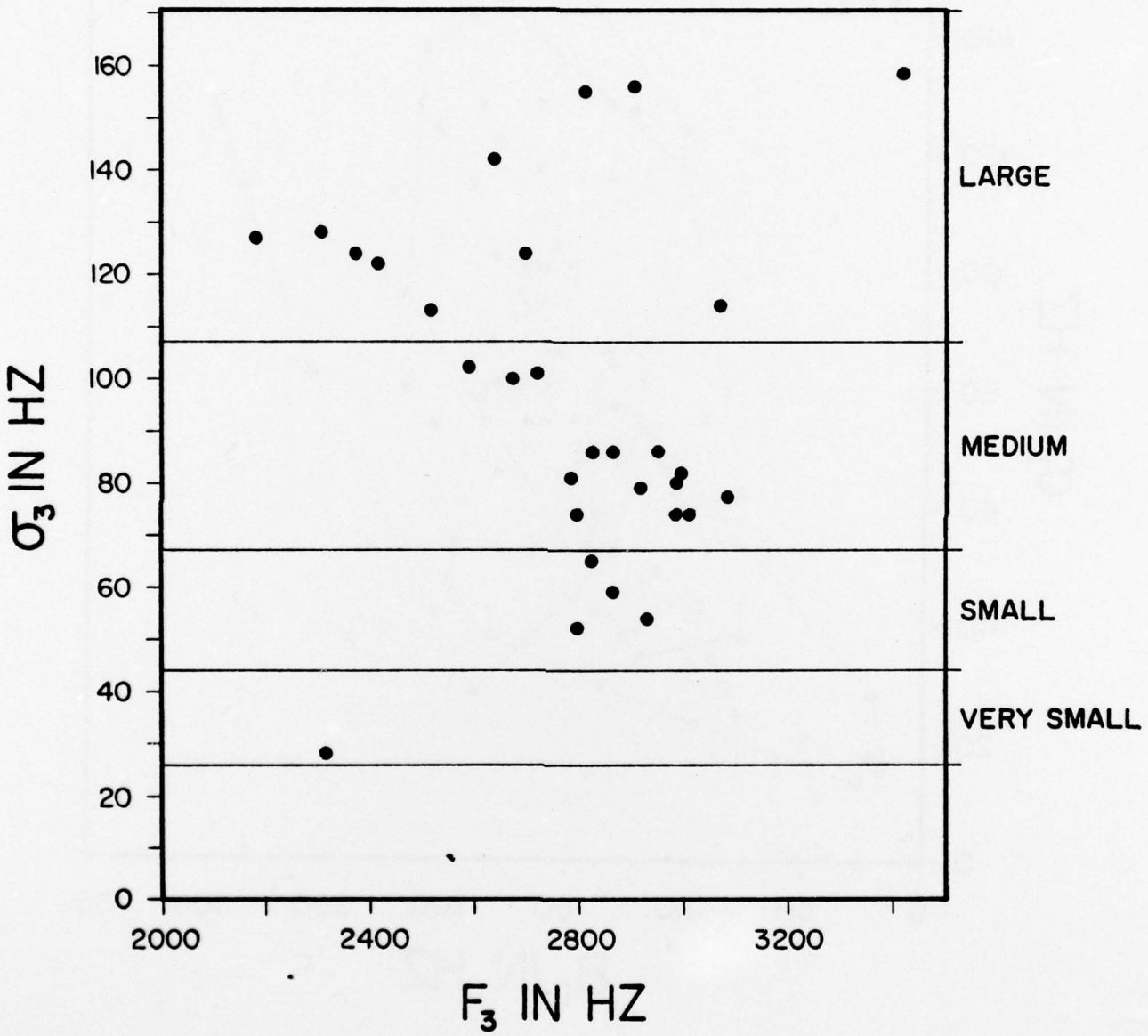


Figure 12. Standard deviation of the third formant frequency plotted against the formant frequency itself.

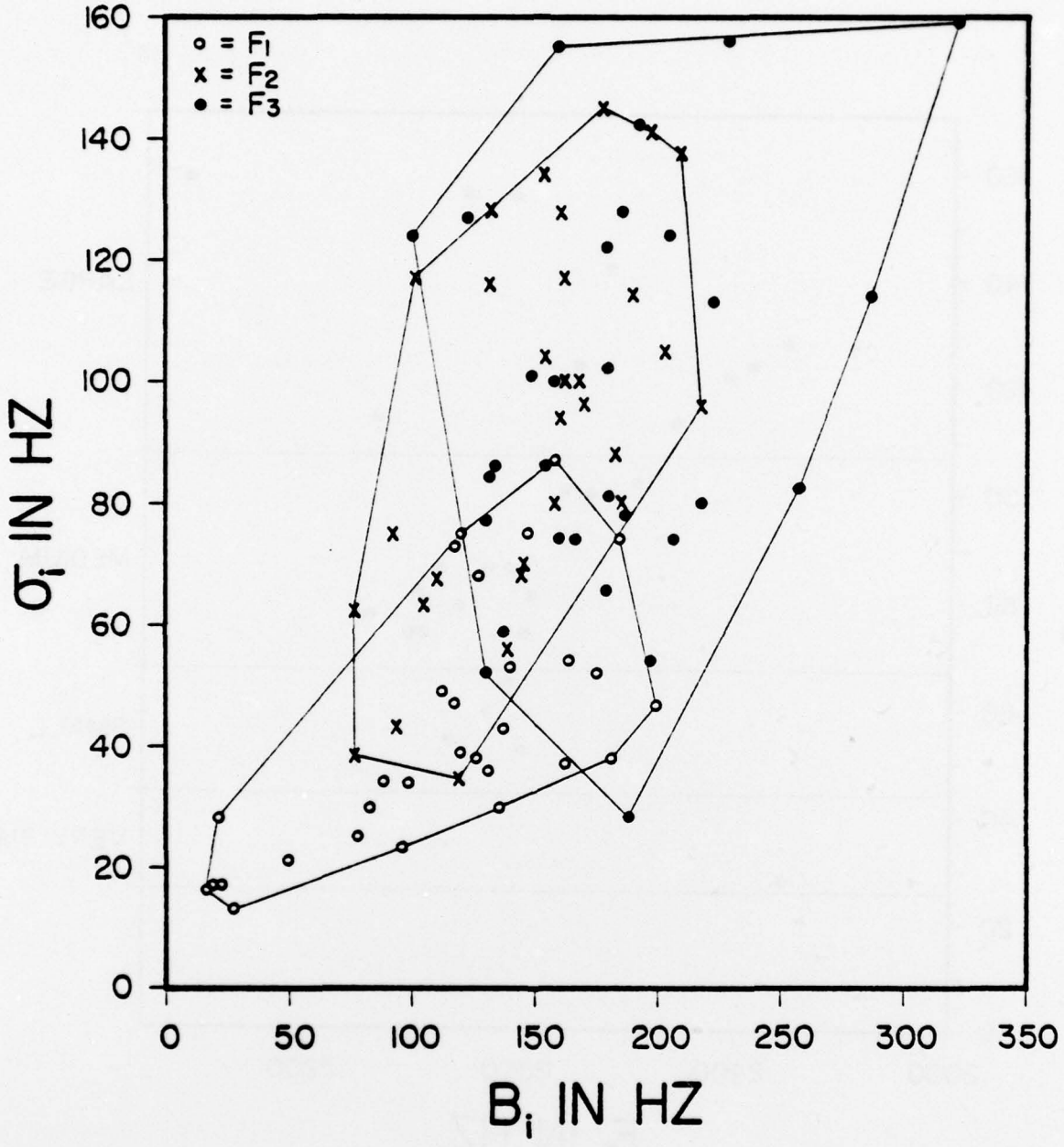


Figure 13. Standard deviations of the formant frequencies plotted against the respective average formant bandwidth estimates. The three loops enclose the data for the three formants.

Implications for Phonetic Categorization

What does this mean in terms of dividing up the formant space into phonetic categories, as suggested at the outset of this paper?

If we were interested only in this one speaker, it would appear that each phonetic category could be defined in terms of its own observed distribution.

But our real goal is to subdivide the acoustic space into catgories that will be general for all speakers. We do not intend here to take up the very popular subject of inter-speaker vowel normalization, but we will assume that such a normalization is possible, so that the coordinates in the acoustic space can be viewed as normalized formant frequencies. Subdividing normalized formant frequencies into categories on the basis of (the size of) the inter-repetition variation would then provide a general classification for vowels produced by any speaker. But this would mean that the categories could not be defined on the basis of data obtained from only one speaker. We will therefore have to look at data from a variety of speakers. Nevertheless, we can use the information from this and preceding studies to form some idea about a general acoustic phonetic categorization for vowels.

We should particularly like to know whether the vowel-to-vowel variation in the inter-repetition variability follows similar patterns in different speakers. The next

slide shows the results obtained in this and other studies [2,6,9] for the vowel [ɪ]. For various reasons, these standard deviations are difficult to compare as an aggregate. Nevertheless, an F_{max} test with 15 degrees of freedom applied to each column reveals no significant difference (among the standard deviations) at the 5% level.

The vowel [ɪ] is one of the more stable configurations in our data. The next slide shows the comparison for the less stable vowel [æ] [6,8,9]. Now the different studies show highly significant differences. Interestingly, the smaller standard deviations shown here for the vowel [ɪ] are comparable to those shown in the preceding slide for [æ]. It therefore seems that vowel-to-vowel differences in variability are not preserved from speaker to speaker. Instead, these limited results suggest an interesting hypothesis, namely, that there is some universal lower limit on the standard deviations for formant frequencies which for any vowel will be approached by some, but not all, speakers.

The present results can be interpreted in terms of this hypothesis to mean that certain of the vowels studied, (especially [ɪ], [ʊ], [y] and [ɯ]) approach this limit on repeatability, while the more variable vowels, (such as [æ] and schwa) do not.

This hypothesis implies that acoustic phonetic vowel categories could be defined as small subdivisions of the normalized formant frequency axes, only a few times as large

VOWEL [ɪ]

SOURCE	STANDARD DEVIATION		
	σ_{F_1} (Hz)	σ_{F_2} (Hz)	σ_{F_3} (Hz)
POTTER & STEINBERG	28	44	43
PETERSON & BARNEY	25	49	—
BROAD & FERTIG	15	55	62
PRESENT STUDY	17	75	77
F_{max}	3.48	2.91	3.21
$F_{max\ crit}$ (5%, 15 d.f.)	4.01	4.01	3.54

⟹ NO SIGNIFICANT DIFFERENCE

Figure 14. Comparison of the present results with those of other studies for the vowel [ɪ].

VOWEL [æ]

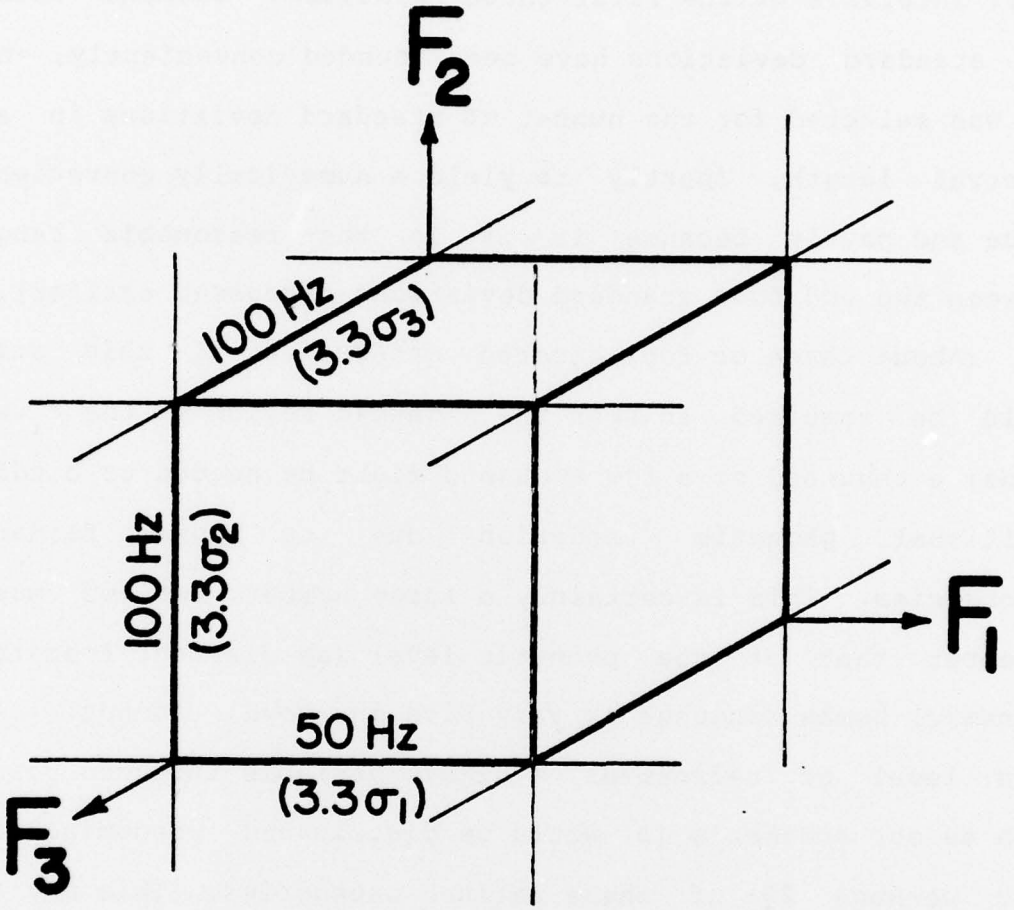
SOURCE	STANDARD DEVIATION		
	σ_{F_1} (Hz)	σ_{F_2} (Hz)	σ_{F_3} (Hz)
POTTER & PETERSON	20	95	—
POTTER & STEINBERG	28	35	43
PETERSON & BARNEY	15	34	—
PRESENT STUDY	75	140	82
F_{max}	25.0	17.0	3.6
$F_{max\ crit}$ (5%, 15 d.f.)	4.01	4.01	2.86

⟹ DIFFERENCES SIGNIFICANT

Figure 15. Same as Figure 14, but for the vowel [æ].

as the smaller observed standard deviations. The last slide shows a hypothetical acoustic phonetic category defined by small intervals of the first three normalized formant axes. The standard deviations have been rounded conveniently, and 3.3 was selected for the number of standard deviations in an interval length, (partly to yield a numerically convenient value and partly because it is in the reasonable range between two and four standard deviations discussed earlier).

(About three or four hundred categories of this size would be required to span the phonetic region of the F_1-F_2 plane; a thousand or a few thousand might be needed to handle additional phonetic variation due to higher formant frequencies. This is certainly a large number, but we must remember that at the phonetic level (as distinct from the phonemic) human language is very rich in vowel sounds. At this level of refinement, a highly variable intended vowel such as our speaker's [æ] would be distributed predominantly over perhaps 20 of these refined categories. This may be somewhat more refined than would be needed for most cases, but it is always easy to define less refined categories as set unions of more refined ones. This is already done within the vowel system of Peterson and Shoup [7], which provides for a possible subdivision of each major vowel category into nine subcategories, thus allowing a choice between a rough level of transcription with 36 categories of place of



$$\text{Min } \sigma_1 \approx 15 \text{ Hz}$$

$$\text{Min } \sigma_2 \approx \text{Min } \sigma_3 \approx 30 \text{ Hz}$$

Figure 16. A tentative acoustic phonetic category in the $F_1F_2F_3$ space defined as a parallelepiped with sides proportional to the smallest observed standard deviations.

articulation and a fine level with 324 such categories. The latter number is about the same as would result from our proposed subdivision of the F_1-F_2 plane.)

There are interesting consequences in defining acoustic phonetic vowel categories in this way [1], but time unfortunately does not permit us to explore this in more detail. The main point is that data on inter-repetition variability can be used to say something about defining the sizes of acoustic phonetic vowel categories.

References

1. D.J. Broad, Toward Defining Acoustic Phonetic Equivalence for Vowels, Phonetica (in press).

2. D.J. Broad and R.H. Fertig, Formant-Frequency Trajectories in Selected CVC English Utterances, The Journal of the Acoustical Society of America, Volume 47, Number 6, Part 2, June 1970, pp. 1572-1582.

3. J.L. Flanagan, A Difference Limen for Vowel Formant Frequencies, The Journal of the Acoustical Society of America, Volume 27, Number 3, May 1955, pp. 613-617.

4. S.E.G. Ohman, Coarticulation in VCV Utterances: Spectrographic Measurements, The Journal of the Acoustical Society of America, Volume 39, Number 1, January 1966, pp. 151-168.

5. G.E. Peterson, The Phonetic Value of Vowels, Language, Volume 27, Number 4, October-December 1951, pp. 541-553.

6. G.E. Peterson and H.L. Barney, Control Methods Used in a Study of the Vowels, The Journal of the Acoustical Society of America, Volume 24, Number 2, March 1952, pp. 175-185.

7. G.E. Peterson and J.E. Shoup, A Physiological Theory of Phonetics, Journal of Speech and Hearing Research, Volume 9, Number 1, March 1966, pp. 5-67.

8. R.K. Potter and G.E. Peterson, The Representation of Vowels and Their Movements, The Journal of the Acoustical Society of America, Volume 20, Number 4, July 1948, pp. 528-535.

9. R.K. Potter and J.C. Steinberg, Toward the Specification of Speech, The Journal of the Acoustical Society of America, Volume 22, Number 6, November 1950, pp. 807-820.

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
Some Statistics on Vowel Formant Variability.

5. TYPE OF REPORT & PERIOD COVERED
Scientific Interim rept.

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
David J. Broad and Hisashi Wakita

8. CONTRACT OR GRANT NUMBER(s)
N00014-76-C-0483 new

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Speech Communications Research Laboratory, Inc.
800A Miramonte Drive
Santa Barbara, CA 93109

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS
Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217

12. REPORT DATE
23 November 1976

13. NUMBER OF PAGES
32

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)
12 32p.

15. SECURITY CLASS. (of this report)
Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)
Distribution of this document is unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES
Paper presented at American Association of Phonetic Sciences, 15 November 1976, San Diego, California

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
formants
Inter-repetition variability
phonetic categories
phonetic equivalence
phonetics
vowels

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)
Acoustic phonetic studies of vowels often present average results with little of the information on formant variability which is essential for defining the sizes of acoustic phonetic categories and for evaluating the comparative significance of contextual and inter-speaker effects. In this study, phonetics instruction tapes provided 828 steady-state tokens of 17 unrounded and 13 rounded non-nasalized, non-retroflexed vowels produced by a single female speaker. The formant frequencies were measured via the linear

20. Abstract (continued)
prediction method, and the most steady-state 108.8 ms of each token was located by an automatic algorithm. Formant analysis errors and wild samples were eliminated by visual inspection of scatter diagrams. The remaining 779 tokens (94.7% of the original tokens) result in standard deviations of between 13 and 87 Hz for F_{1}^{SUB1}, 32 and 145 Hz for F_{2}^{SUB2}, and 28 and 159 Hz for F_{3}^{SUB3}. Except for some indications of multi-modality for cases of very high standard deviation, no discernable systematic relation seems to exist between formant variability and (1) articulatory features of the vowels, (2) nativeness of the vowels, (3) formant frequency, or (4) formant bandwidth. Although some of the standard deviations are larger than expected on the basis of earlier studies, certain vowel configurations are found to be highly reproducable. This finding is related to acoustic phonetic vowel classification.